

## Mental Workload and Visual Impairment: Differences between Pupil, Blink, and Subjective Rating

Miguel Ángel Recarte<sup>1</sup>, Elisa Pérez<sup>1</sup>, Ángela Conchillo<sup>1</sup>, and Luis Miguel Nunes<sup>2</sup>

<sup>1</sup>Universidad Complutense (Spain)

<sup>2</sup>Dirección General de Tráfico (Spain)

This research has two aims: (a) To study the concurrent validity of three measures of mental workload, NASA TLX rating scale, pupil dilation and blink rate, testing the hypothesis that they will provide convergent results using a single-task, and dissociative results for dual-task; and (b) To analyse their capability to predict visual search impairment. These three measures were analyzed in the same cognitive tasks in single-task and dual-task (cognitive task and visual search) conditions in a within-subjects experiment with twenty-nine participants. Mental workload measures showed concurrent validity under single-task condition, but a complex pattern of results arose in the dual-task condition: it is suggested that NASA TLX would be a subjective addition of the rating of each task; pupil dilation would measure the average arousal underlying the cognitive tasks; and the blink rate would produce opposite effects: whereas mental workload of cognitive tasks would increase blink rate, visual demand would inhibit it. All three measures were good predictors of visual impairment. The soundness of these measures is discussed with regard to the applied field of driving and other activities.

*Keywords:* mental workload, pupil, blink, NASA TLX rating scale, distraction in driving

Este experimento tiene dos objetivos: 1) Estudiar la validez concurrente de tres medidas de carga mental, la escala de juicios NASA TLX, la dilatación de la pupila y la tasa de parpadeo, poniendo a prueba la hipótesis de que, en situaciones de tarea única, arrojan resultados convergentes, pero, en doble tarea, arrojan resultados disociativos. 2) Analizar su capacidad para predecir el deterioro en la búsqueda visual. Las tres medidas fueron analizadas con las mismas tareas cognitivas realizadas en condiciones de tarea simple y de doble tarea (tarea cognitiva y búsqueda visual) en un experimento intrasujetos con veintinueve participantes. Las medidas de carga mental mostraron validez concurrente en las condiciones de tarea única, pero en las condiciones de doble tarea apareció un patrón de resultados complejo que sugiere que NASA TLX consistiría en la adición subjetiva de los juicios de cada tarea; la dilatación de la pupila mediría la activación promedio que subyace a las tareas cognitivas; y la tasa de parpadeo produciría efectos contrapuestos: mientras que la carga mental de las tareas cognitivas incrementa la tasa de parpadeo, las demandas visuales la inhiben. Las tres medidas fueron buenos predictores del deterioro visual. Se discute la justificación del uso de estas medidas en el campo aplicado de la conducción y otras actividades.

*Palabras clave:* carga mental, pupila, parpadeo, escala NASA TLX, distracción al conducir

---

Correspondence concerning this article should be addressed to Miguel Ángel Recarte, Facultad de Psicología, Universidad Complutense de Madrid, Campus de Somosaguas, 28223-Madrid, Spain. E-mail: [marecarte@psi.ucm.es](mailto:marecarte@psi.ucm.es)

How to cite the authors of this article: Recarte, M.A., Perez, E., Conchillo, A. and Nunes, L.M.

The concept of mental workload presents a multiplicity of psychological approaches: some rely on information processing stages (De Waard, 1996, 2002), others on the concept of effort (Vicente, Thornton, & Moray, 1987), and others on the limited capacity of our processing system (O'Donnell & Eggemeier, 1986). Despite these differences, there is a general agreement: mental workload is seen as the result of an interaction between task demands and human characteristics, that is, separately, neither the task properties nor the human operator characteristics can explain mental workload. In order to overcome the lack of an operative definition, efforts have been made in the last few years to find measures of mental workload, given the usefulness of such a concept to explain human performance errors in terms of operator overload. This is important, particularly when the knowledge and prediction of human information processing errors becomes crucial to improve human interaction with systems involving risk, including accident prevention in various contexts. Examples of such contexts are training of air traffic control operators (Wickens, 1992), and drivers' performance in road traffic (De Waard, 1996; De Waard & Brookhuis, 1997).

There has been a remarkable increase of scientific contributions addressing attentional aspects in road traffic during the last years, and the present study is derived from the following group of previous studies on mental load and visual behaviour during driving: Recarte and Nunes (2000) studied the effects of mental workload induced by cognitive tasks on eye movements and visual search behaviour. The study was replicated and extended to show some effects of mental workload on visual detection and discrimination capacities (Recarte & Nunes, 2003). The same authors studied attentional load issues related to speed control, and the mental load associated with mobile phone versus live conversation with a passenger while driving (Nunes & Recarte, 2004; Recarte & Nunes, 2002a). The effects of mental workload and time spent driving on blink rate were also studied in Recarte and Nunes (2002b). These experiments, all of them performed in real traffic while driving a car, provided with an eye-tracking-system, inspired the present study in an attempt to consolidate, in a controlled laboratory environment, some of the previously obtained results on the road, and to verify some hypotheses about the concept of mental workload. The selection of independent variables (cognitive tasks and visual demand), mental workload measures, and the design of the present experiment were also inspired in these cited researches on the road.

The purpose of the present experiment is to compare three different measures of mental workload (NASA-TLX scale, pupil dilation and blink rate), taken while performing the same cognitive tasks with/without a visual search task. As a general hypothesis, it is assumed that if the three measures are measuring the same construct of mental workload, they should behave similarly in the cognitive tasks and all three should produce deterioration in the visual

search task. Hits in detection are included as external criteria to determine whether the construct of mental workload, that is implicit in the measures, explains human error in visual perception performance. Specific hypotheses about each workload measure are expounded below.

### *The workload measures*

*The subjective rating scale: NASA-TLX (Task Load Index).* Some authors consider that subjective report is the best way to evaluate mental workload, with the advantage that individual differences are already included in such measures (De Waard, 1996). In fact, subjective ratings have been used to validate physiological measures (Iqbal, Zheng, & Bailey, 2004). However, several internal processes are far from being accessible to consciousness, or people may not discriminate between different task dimensions (such as difficulty vs. effort), or, if they do, different values could be reported because of different decision criteria. De Waard, while accepting that humans can be aware of mental workload in the case of single tasks, points out that this is hardly possible in dual-task contexts. Our experiment addresses this issue by comparing mental workload rating in dual-task versus single-task performance.

In the present study, the NASA-TLX (Hart & Staveland, 1988) was chosen as the subjective measure. It is a multidimensional scale with six subscales: mental demand, physical demand, temporal demand, effort, performance, and frustration. It is administered in two steps: in the first step, the participants evaluate, by means of pair comparison and before knowing the specific tasks to be evaluated, the relative contribution of each of the six dimensions represented in each subscale to the total evaluation of workload. The scores thus obtained are used in the second step to attribute a weight to each subscale in the final score. In the second step, the participant rates the task performed on each subscale (ranging from 0 to 20) and, according to the previously obtained weight factors, the final score is obtained.

According to Rubio, Díaz, Martín, and Puente (2004), the NASA-TLX has showed to be sensitive and has worked successfully in a broad array of tasks. The same authors noted the advantages of the multidimensionality of the scale, providing specific information about different sources of mental workload, scarce intrusion in task performance, and predictive validity of performance in various tasks.

*Pupil dilation.* Pupil dilation has been considered a reliable measure of mental workload (Beatty, 1982; Hoecks & Levelt, 1993; Janisse, 1977; Kahneman, 1973). But, in addition to pupil response to mental workload, pupil dilation can be biased due to light conditions, emotional responses, or orienting reflex (Näätänen, 1992). According to Granholm and Steinhauer (2004), pupil dilation reflects activation of brain modules, no matter whether the information processed

is neutral or of an emotional nature. Using fMRI, Just, Carpenter, Keller, Emery, Zajac, and Thulborn, (2001) found, in the case of dual-task (verbal and spatial), that the total brain activation in both activated areas (parietal and temporal) is much lower than the sum of the activation induced by each task when performed alone, which is consistent with the existence of upper limits for brain activation (or for some brain areas). This is in line with the concept of attention from a limited resources perspective and would explain impairment due to dual-task performance. Assuming that, hypothetically, pupil dilation reflects the whole brain activation, in dual task conditions, we should expect that the pupil dilation effect would be lower than the sum of the effects of each task when performed alone.

In previous studies, Recarte and Nunes (2000) showed that pupil size was sensitive to increased mental workload induced by secondary cognitive tasks during driving, and this effect could be clearly shown despite the variability due to daylight variations on a natural road environment. The same results were replicated in Recarte and Nunes (2003), and the pupil data seemed to be consistent with subjective effort ratings, although such ratings were set in a unidimensional scale. But in the driving context, it cannot be ruled out that, besides the cognitive effort of dual tasks, there may be emotional components that increase because of the awareness of risk. These results encouraged us to carry out a more controlled laboratory study, in which the pupil effect of the same cognitive tasks would be compared with a more refined subjective multidimensional scale (NASA-TLX) and the possible emotional components associated with real driving could be eliminated.

As in the above-mentioned driving experiment, the cognitive tasks would also be measured under a dual-task paradigm, and this would allow us to test two hypothetical assumptions. The first one deals with the specificity of the pupil effect: the pupil dilation would reflect the purely mental component of the total workload imposed by the task. In the driving context, the pupil dilation effect can be observed when mentally anticipating complex manoeuvres such as overtaking or approaching a roundabout, that is to say, during mental manoeuvre planning. The second hypothesis deals with the expected effects under dual-task condition: the pupil would not reflect the sum of effects of each task when performed alone, but some value between the two, maybe an average value or a maximum value that could reflect an upper limit of activation of some brain area.

*Blinks.* Although De Waard (1996) proposes blink rate as a mental workload measure, his references do not actually correspond to the strict concept of cognitive load but to other aspects of the task: when citing Stern, Boyer, and Schroeder (1994), these authors, in fact, concluded that the increase in blink rate is a consistent indicator of fatigue, and this result led to several attempts to design in-vehicle safety devices to detect drivers' fatigue and drowsiness states.

According to Fogarty and Stern (1989), a reduced blink rate reflects the increased visual demands of a task, as a simple mechanism to reduce the probability of missing relevant information. Siveraag and Stern (2000) also point out that the blink inhibition effect corresponds to more demanding tasks. Although these authors do not explicitly mention the visual nature of the tasks, all their examples refer to visual tasks. But what can be expected from cognitive tasks with no visual component, such as calculating or talking?

Ryu and Myung (2005) use the time between blinks, together with two measures derived from EEG and heart rate, as measures of mental workload in dual-task conditions: a tracking task and mental arithmetic. The authors report that the interval between blinks allows inferring about the tracking task load but not about the mental arithmetic task. They conclude that all three measures are needed to evaluate the workload of the different tasks.

Recarte and Nunes (2002b) found a blink-rate increase attributable to the time spent driving, but they also found an even more marked increase attributable to the performance of cognitive tasks during the drive. This poses an interpretation problem: long time driving leads to a decreased activation state with an increased blink rate, and the performance of a cognitive task leads to an activation response (according to pupil dilation) but also to an increased blink-rate effect. That is to say, a blink-rate increase can be expected either from fatigue (decreased activation) or from the mental effort of a secondary task (increased activation). The authors, based on suggestions from Stern et al. (1994), hypothesized that the strong visual component of the driving task could lead to a general blink inhibition effect. Such an effect could be interpreted as a continuous inhibitory control task. The blink-rate increase associated with long time driving (fatigue) could be interpreted as impairment of such inhibitory control. In contrast, the performance of a secondary cognitive task could have an analogous effect: the increase of resources needed to perform the mental task while driving would interfere with the resources needed for blink inhibition. Such an interpretation, although speculative, could explain why blink increase can be expected either from fatigue or from additional cognitive effort.

In our experiment, there was baseline condition consisting of looking at a screen where different letters appeared at a rapid rate. We expected a blink-rate increase, with regard to baseline, when performing a cognitive task without visual demand, and this increment would be related to the increment of mental workload of the task. But if the task is more visually demanding, such as searching and detecting targets on the screen, a lower blink rate would occur, compared to the baseline condition. Finally, in this visual search condition, we hypothesized that the performance of a secondary cognitive task would produce a blink-rate increase. Summarizing, there were two basic

hypotheses: a) tasks requiring continuous visual search would produce blink-rate reduction; b) cognitive tasks with no visual component, if performed together with the continuous visual search task, would produce a blink-rate increase, depending on the mental workload of the cognitive task.

### *The Cognitive Tasks Inducing Mental Workload*

The experiments of Recarte and Nunes (2003) included actual driving in real traffic while performing listening tasks the content of which the participants were requested to attend, as they would subsequently be asked to reproduce such content in their own words. In addition, participants also had to simultaneously perform a visual task (detection and discrimination of flashing lights). This task was used as performance criterion to evaluate possible visual impairment due to mental workload. By comparing the effects of listening versus talking, the authors found low subjective effort rates and scarce or null effects on pupil size for the listening, and high subjective rates and significant pupil-size increase for the talking tasks. These effects were interpreted as resulting from the increased mental load of the talking tasks. The effects on the visual discrimination task were in accordance with the results on mental workload: no effects due to listening tasks, and impairment of visual stimulus processing (lower detection rate and more decision errors) with the talking tasks. In this experiment, subjective effort rating, pupil-size effect and visual performance measures varied accordingly. In addition to the discussion of its applied consequences in the driving context, we also note that Kubose, Bock, Dell, Garnsey, Kramer, and Mayhugh (2006) found no differences between comprehension and verbal production in conditions of simulated driving, using several driving performance parameters as dependent measures.

Consequently, we considered of particular interest the replication of previous results of Recarte and Nunes (2003) in a laboratory setting, including the same listening and talking tasks, and a third cognitive task, comparable to the talking in terms of mental workload, but allowing a more objective performance evaluation: a mental arithmetic task, (counting down by successively subtracting a given amount, and starting from a given number). This task has been used in the above-mentioned driving experiments, and also to evaluate the load of the central processor of the working memory (HASTE, EU FP5 project, 2002), and in other experiments on mental load and pupil size (Steinhauer, Siegle, Condray, & Pless, 2004).

### *The Visual Search Task*

Finally, a continuous visual search task is the key for the study of mental workload measures. According to our starting point, this task implies a high visual demand

(detecting a target letter among 16 letters) and low mental effort (to identify the target letter and to press a key). More details are provided in the Method / Procedure section. In consequence, we should expect low pupillary dilation and blink rate reduction and a medium value in NASA-TLX score, attributable to the time pressure imposed by the task. In Recarte and Nunes (2003), the visual task consisted of detecting discrete luminous stimuli while driving, and either performing or not performing a mental task. It is possible that part of the deterioration in detection had to do with the use of the glances necessary to drive. This is another reason for repeating these results in laboratory conditions.

This task, combined with the above-mentioned cognitive tasks, allows introducing the dual-task paradigm in the design. While in the single-task condition, the three above-mentioned cognitive tasks (verbal acquisition or listening, verbal production or talking, and mental arithmetic or calculating) are performed alone, in the dual-task condition, equivalent versions of the same tasks are performed simultaneously with the visual search task. In both cases, the three mental workload measures are obtained.

With regard to the comparison between single versus dual-task, our expectations are that, in the dual-task condition, the NASA-TLX score would probably reflect the sum of the effort of both tasks, the pupil would reflect only the cognitive component (not the increased effort merely associated with performing two tasks), and blink rate would split the visual component from the cognitive one: the higher visual demand would produce lower blink rate in the dual-task compared to the single-task, and the cognitive component would produce an increased blink rate compared to the previously observed reduction attributable to the visual search.

Summarizing, the present study attempts to clarify the concurrent validity of the three above-mentioned measures of mental workload (NASA-TLX, pupil, and blink), using experimental conditions of single-task (characterized by performing some cognitive tasks with mental workload, but without visual demand) and some other dual-task (the former cognitive tasks conjointly with a visual search task with a strong visual demand). The general hypothesis states that, whereas it is expected that the workload measures will converge in their results in the single-task conditions, they will not reflect a consistent pattern when attempting to measure the mental workload implied in the dual-task conditions: considering the above-mentioned hypothesis about pupil dilation, we expect to test its specificity as an indicator of mental workload and its non-additive effect in the dual-task conditions. With regard to blink rate, we attempt to provide evidence for the disinhibition hypothesis: the increase of blink rate with mental workload of the cognitive task is interpreted as the disinhibition of the blink inhibition induced by the strong visual demand required by the visual search task.

## Method

### Participants

The study was carried out with a sample of 29 volunteer students from the Psychology Faculty of the Universidad Complutense de Madrid. Their ages ranged from 18 to 23 years.

### Experimental Design

The independent variables are as follows:

1. Cognitive task, with four alternatives: listening, talking, calculating, and a control condition (no cognitive task).
2. Visual demand, with two levels: visual search or detecting (detecting a target letter), and no visual search (no detecting).

By combining Cognitive Task x Visual Demand, we obtain eight experimental conditions in a within-subject design: four conditions without detection (listening, talking, calculating, and control), and four of them with detection (listening & detecting, talking & detecting, calculating & detecting, and control & detecting). Each experimental condition lasted for two minutes.

The three experimental conditions, listening & detecting, talking & detecting, and calculating & detecting, will be called the *dual-task condition*, because they require performing at the same time a cognitive task and the visual search task. Similarly, listening, talking, and calculating (without detecting) will be called the *single-task condition*, because these tasks do not require visual effort. In the conditions of no visual search (no detecting), the subject only had to look at the screen (without detecting target letters) in order to match the visual environmental conditions (lightning and type of stimuli), and to make the single-task conditions comparable to the dual-task conditions.

The talking conditions always had to be performed immediately after the listening conditions, because talking consisted of producing a verbal report of what the participants had just listened to.

All the conditions with visual search were performed after completing the conditions without visual search. This

was decided because, once the participants were given the instruction to detect target letters for the first time, it would have been unfeasible to erase the subjective expectations for target searching and to recover the passive looking that corresponded to no visual search conditions. Consequently, all the dual-task conditions were performed after the single-task ones.

### Materials and Procedure

Materials were eye-tracking system (ASL 5000), video recorder, screen (146x110 cm), and NASA-TLX scale.

Visual stimuli consisted of capital letters that were projected onto a rectangular screen, and participants were seated so that the screen covered a visual angle of 35° horizontally by 27° vertically. In order to optimize the significance of the pupil data, the luminance of the screen was kept at a constant level of 110 lux. The participants' head position was kept stable by means of the shape of an anatomic chair, with no need of firm fixation such as a chinrest, and no equipment attached to their heads. This was possible due to the eye-tracking system used (ASL 5000), which provided a 50 Hz measurement of eye-gaze coordinates and pupil size. A computer program presented the general and specific instructions for each condition to the participants, allowing a self-paced rhythm to perform the sequence of the eight experimental conditions, and it also recorded their responses. A video camera recorded the eyes and the verbal responses.

The experimental sequence was as follows:

1. NASA-TLX, first part: According to the instructions of this measure, the participants had to compare pairs of the six NASA-TLX dimensions: mental demand, physical demand, time pressure, effort, performance, and frustration, and then evaluate which dimension within each pair is closer to the concept of workload. These results were used to weight the results of the scale in the experimental phase.
2. Calibration of the eye-tracking system and performance of the eight experimental conditions, the first four without visual search, and the next four with visual search. The second part of NASA TLX was completed after each condition.

In each condition, the participants observed a frame

Table 1

Cognitive Task				
Visual demand	No (Control)	Listening	Talking	Calculating
Visual Search	Control & Detecting	Listening & Detecting	Talking & Detecting	Calculating & Detecting
No Visual Search	Control & No detecting	Listening	Talking	Calculating



containing 16 capital letters that were simultaneously projected onto the screen. These letters changed asynchronously: every 125 ms, a letter was substituted by a different one and each letter was present for 2s. The location of each change and the letter that appeared were randomized, although the same letter could not appear in two places at the same time. In the visual search conditions, the participants had to detect the letter R.

During the entire experiment, the pupil diameter, the gaze's coordinates, the blinks, the time at which the stimulus appeared, and the subject's response, were registered.

## Results

A repeated measures ( $4 \times 2$ ) ANOVA (Cognitive Task with four conditions  $\times$  Visual Demand with two conditions) was run (SPSS 14). In addition, contrasts to the control conditions and binary comparisons between cognitive tasks were obtained, controlling error type I by Bonferroni's method. Information about the effect size and power is provided. The results of the ANOVAs can be seen in Table 2. Means and standard errors of the mental workload measures and proportion of the detection hits are presented below in figures.

Table 2

*Within-Subject ANOVA, and Simple Contrasts of the Cognitive Tasks Compared to the Control Condition for the NASA TLX, Pupil Size, Blink Rate, and Detection Hits*

Measure	Variation Source	Cognitive Task	<i>df</i>	<i>F</i>	<i>p</i>	$\eta^2$	<i>Power</i>
NASA TLX	Visual Demand		1, 28	36.52	.000	.566	1.00
	Cognitive Task		3, 84	43.09	.000	.606	1.00
	Simple Contrast	Listening	1, 28	15.31	.001	.354	.965
		Talking	1, 28	63.03	.000	.692	1.00
		Calculating	1, 28	62.78	.000	.692	1.00
	Visual Demand $\times$ Cognitive Task		3, 84	5.02	.003	.152	.904
	Simple Contrast	Listening	1, 28	0.60	.444	.021	0.117
		Talking	1, 28	1.48	.234	.050	0.227
		Calculating	1, 28	5.94	.021	.175	0.653
Pupil Size	Visual Demand		1, 28	5.13	.031	.155	.589
	Cognitive Task		3, 84	78.93	.000	.738	1.00
	Simple Contrast	Listening	1, 28	3.95	.057	.124	.484
		Talking	1, 28	84.93	.000	.752	1.00
		Calculating	1, 28	81.93	.000	.745	1.00
	Visual Demand $\times$ Cognitive Task		3, 84	51.49	.000	.648	1.00
	Simple Contrast	Listening	1, 28	6.23	.019	.182	.674
		Talking	1, 28	89.42	.000	.762	1.00
		Calculating	1, 28	74.13	.000	.726	1.00
Blink Rate	Visual Demand		1, 28	64.23	.000	.696	1.00
	Cognitive Task		3, 84	42.66	.000	.604	1.00
	Simple Contrast	Listening	1, 28	42.71	.000	.604	1.00
		Talking	1, 28	62.86	.000	.693	1.00
		Calculating	1, 28	65.51	.000	.701	1.00
	Visual Demand $\times$ Cognitive Task		3, 84	4.01	.010	.125	.822
	Simple Contrast	Listening	1, 28	5.41	.028	.162	.612
		Talking	1, 28	15.86	.000	.362	.970
		Calculating	1, 28	6.70	.015	.193	.705
Detection Hits	Cognitive Task		3, 84	41.91	.000	.599	1.00
	Simple Contrast	Listening	1, 28	9.00	.006	.243	.825
		Talking	1, 28	60.08	.000	.682	1.00
		Calculating	1, 28	72.22	.000	.721	1.00

### NASA-TLX

Data on workload rating of NASA TLX scale are displayed in Figure 1.

A statistically significant effect of visual demand indicates that participants judged that detecting has more workload than no detecting. That is, they evaluated that performing a cognitive task while detecting visual targets has more workload than just looking without detecting. At the level of the control condition, the effect of visual demand indicates that visual detection presents more workload than just looking at the visual stimuli without detecting.

The variable cognitive task produced significant differences in the participants' ratings, and also when comparing each task with the control condition. Although not displayed in Table 2, pair-wise comparisons indicate the existence of significant differences between cognitive tasks, with the exception of the pair talking and calculating.

The interaction showed a significant but low effect, indicating that the difference observed in calculating between the single (only calculating) and dual-task (calculating & detecting) conditions is lower than the differences observed in the other cognitive tasks between single and dual-task conditions, and also for the control condition. As this effect is weak and only present for calculating, it is difficult to interpret.

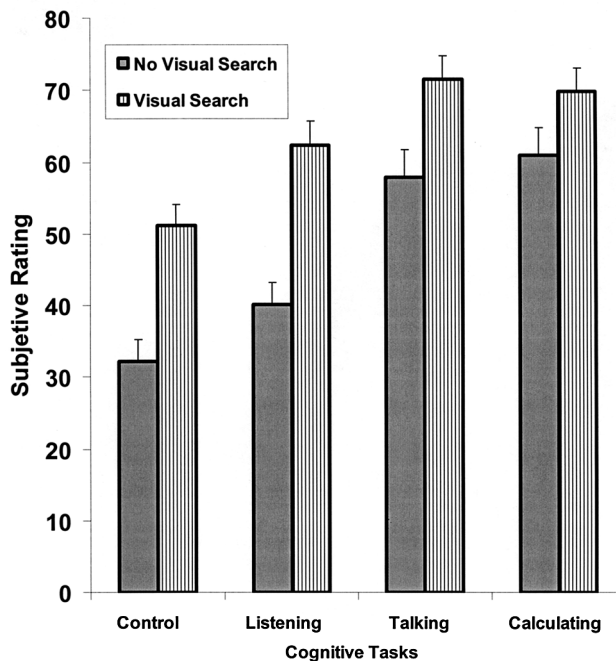


Figure 1. Mean and Standard Error of NASA TLX by Visual Demand and Cognitive Task.

### Pupil Size

The results indicated a statistically significant effect of visual demand (see Table 2 and Figure 2): pupil size was larger when performing cognitive task and detecting (dual-task) than when only performing the cognitive task without detecting (single-task). This effect was smaller than the one observed in the subjective rating scale NASA-TLX, and it was shaded by interaction. In the control condition, pupil size was larger when detecting than when merely looking without detecting.

Statistically significant differences appeared in the variable cognitive task, and, with the exception of listening, the comparisons of each task with control were also significant. Significant differences also appeared in the pairs listening-talking and listening-calculating, but not in the case of talking-calculating (as in NASA TLX).

A significant interaction between cognitive task and visual demand indicated that pupil dilation response is larger in dual-task (cognitive task with detection) than in single-task (cognitive task without detection) for cognitive tasks with lower mental workload (control and listening), whereas in the case of the tasks with higher mental workload (talking and calculating), there was a higher pupil response in the single-task compared to dual-task. This interaction is also significant for each of the three cognitive tasks with respect to the control condition.

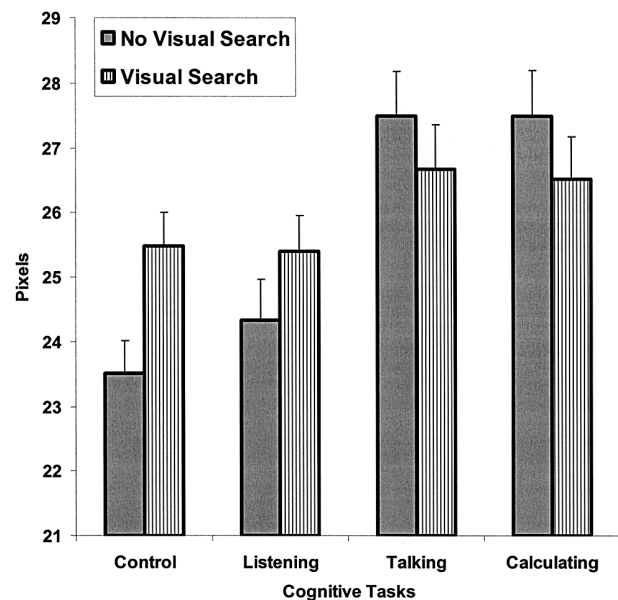


Figure 2. Mean and Standard Error of Pupil Size (pixels) by Visual Demand and Cognitive Task.

### Blink Rate

The results can be seen in Table 2 and Figure 3.

Both the statistically significant effects attributable to visual demand and those attributable to cognitive task corroborated the above-mentioned hypothesis: dual-task conditions (cognitive task with detecting) appeared associated with lower blink rate. In addition, all the cognitive tasks produced blink-rate increase when compared to the control condition, and pair comparisons showed differences between tasks, except for talking and calculating.

The interaction Cognitive Task x Visual Demand was statistically significant, and the contrast of this interaction was also significant for all levels of the cognitive task compared to the control condition, indicating that blink-rate difference between no detecting and detecting is lower in the control condition (no cognitive task) than for the three conditions of cognitive task (listening, talking, calculating).

Therefore, blink rate appears to be as sensitive as NASA-TLX or pupil size with regard to mental workload, but moreover, it clearly discriminates between visual demand and mental workload with opposite effects: the former leading to blink-rate inhibition and the latter to blink-rate increase.

For the three measures, NASA, pupil and blink, in table 1 we can see the effect size and the power. Regarding Visual Demand, for NASA-TLX ( $\eta^2 = .566$ , power = 1) and Blink rate ( $\eta^2 = .696$ , power = 1), the effect size and power are

higher than for pupil ( $\eta^2 = .155$ , power = .589), indicating that the pupil does not reflect the total workload attributable to dual-task (cognitive task & detecting). On the other hand, we note that for all measures (NASA-TLX, pupil, blink rate, and proportion of hits), the effect sizes attributable to the cognitive task are near  $\eta^2 = .60$ , with the pupil effect size being the highest one ( $\eta^2 = .738$ ), and also power was equal to 1 in all of them, which indicates the importance and the robustness of differences found.

### Proportion of Detection Hits in Visual Search Condition

Here, visual performance is analysed in terms of hit rate (correct detection responses) with respect to the diverse cognitive tasks. Figure 4 shows the mean proportion of hits as function of the cognitive task when performed together with the detection task.

In general terms, the hit-rate percentage reduction due to cognitive tasks follows the pattern of mental work load measure variation: a variation of 8 points, between control (just detecting) and detecting & listening, a marked reduction of 16 points between detecting & listening and detecting & talking, and finally, a lower difference of 9 points between detecting & talking and detecting & calculating.

In the ANOVA, the variable cognitive task showed significant differences in the proportion of detected stimuli,

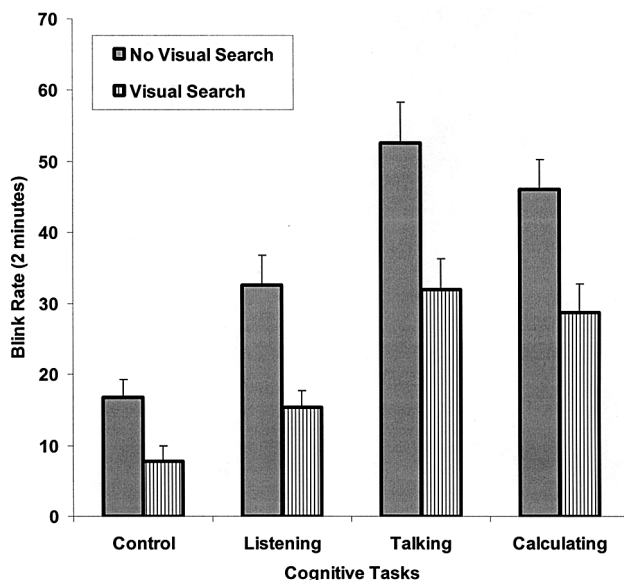


Figure 3. Mean and Standard Error of Blink Rate by Visual Demand and Cognitive Task.

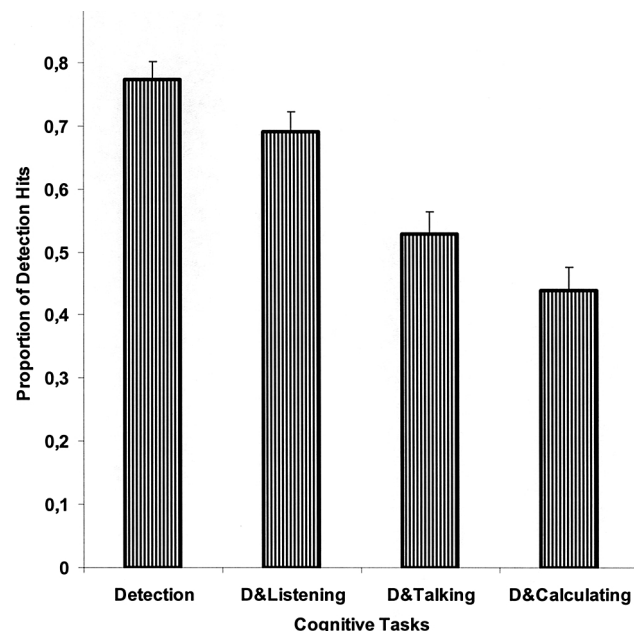


Figure 4. Proportion and Standard Error of Detection Hits by Cognitive Task.



$F(3, 84) = 41.91, p < .001$ . Statistically significant differences appeared between the control condition and the three dual-task conditions. Also, all the binary comparisons, except for detecting & talking versus detecting & calculating, yielded statistically significant differences.

### Correlations

The figures presented above suggest high correlations between mental workload measures across cognitive tasks and negative correlations with visual detection. They also suggest that different relations appear when cognitive tasks are performed alone or in dual-task conditions. Therefore, we analyzed these correlations separately for single-task and for dual-task conditions: correlations between mental workload measures were computed, and also between hits in detection and mental workload measures in the dual task.

*Correlations of mental workload measures with hit rate.* Table 3 displays the results of the above-mentioned correlations.

All the correlations were statistically significant,  $p < .01$ .

*Correlations between the three mental workload measures.* When comparing the correlations between mental workload measures with regard to both conditions of visual demand (visual search vs. no visual search), the data show lower correlations for visual search (dual-task) conditions. Data can be seen in Table 3.

In the case of NASA-TLX and pupil, the correlation drops drastically from .77 in single-task to .42 in dual-task. Although under single-task conditions, one could state that NASA-TLX and pupil measure the same mental workload construct, this assertion lacks support in the case of dual-task. This difference in correlations is in accordance with the above-reported experimental data: a similar pattern for both measures in single-task but not

in dual-task, where pupil presents an interaction, consisting of a lower pupil size for tasks with higher mental workload and vice versa, but NASA TLX does not reflect this interaction effect.

### Discussion

In general terms, the results confirm our expectations: the three mental workload measures —NASA-TLX, pupil dilation, and blink rate— show concurrent validity when evaluating the three cognitive tasks (when performed alone), allowing the inference of a basic underlying construct of mental workload. Data from ANOVA and from correlational analysis agree, for the three measures, with the following statements: (a) all the cognitive tasks produce a higher mental workload than control condition (no task), (b) talking and calculating produce a higher mental workload compared with the scarce mental workload of listening, and (c) previous results of Recarte and Nunes (2003), in a driving context, are replicated with regard to differences between listening and talking, and listening and calculating.

Considering the workload imposed by the visual detection task, the data also confirm our expectations regarding the particularities of blink rate variations as discriminating between visual and mental workload. According to NASA-TLX scores or pupil data, visual and mental workload can be set according to a unidimensional measure of workload. Detecting is slightly more loading than just looking or listening, but lower than talking or calculating. According to blink rate, visual and mental workload produce opposite effects: Blink inhibition for higher visual demand and increased blink rate for higher mental workload. Therefore, NASA-TLX and pupil do not differentiate between visual and mental workload, but blink does.

Table 3

Pearson Correlations between the Mental Load Measures and Hit Rate in Detection Dual-Task (Cognitive Task & Detection), and Single-Task (only Cognitive Task) Conditions

Measures	Single-Task	Dual-Task
NASA TLX – Hits	-.61**	-.62**
Pupil Size – Hits	-.69**	-.48**
Blink Rate – Hits	-.60**	-.52**

Table 4

Correlation between the three Mental Load Measures for Dual-Task (Cognitive Task & Detection) and Single-Task (only Cognitive Task) Conditions

Measures	Single-Task	Dual-Task
NASA TLX – Pupil Size	.77**	.42**
NASA TLX – Blink Rate	.60**	.57**
Pupil Size – Blink Rate	.72**	.45**

For single task compared with dual-task, NASA-TLX displays an additive pattern and blink a subtractive one, whereas the case of pupil shows markedly different effects when evaluating the same tasks, depending on single versus dual conditions. For NASA TLX, the total workload appears to be the sum of the workload of each task (mental plus visual), and blink rate is the result of two additive effects although with different sign (the more mental workload, the more blinks; the more visual demand, the less blinks). The interaction of pupil response indicates that, in single-task conditions, there is a higher pupil-dilation effect of the more loading mental tasks than in dual-task, and vice versa for less loading mental tasks. We suggest that the pupil could reflect the highest of the activation states or perhaps an average value of the brain activation areas associated with the task performance. Hypothetically, areas associated with the tasks with higher mental workload (talking and calculating) would be less activated in dual-task conditions because, in this case, a part of such activation would correspond to resources shared with the detection task.

Regarding the prediction of impairment in a visual search task, we expected that mental workload measures would act as predictor variables, although different results in mental workload evaluation when comparing single-task and dual-task were also obtained: visual performance impairment is better predicted from single-task workload evaluation than from an evaluation carried out in dual-task conditions.

According to the results, the equivalence of mental workload measures becomes a more complex subject under dual-task conditions. Subjective evaluation results (NASA-TLX) suggest that our judgements of total mental workload, attributable to both tasks (mental task and visual search), seem to be based more on some additive reconstruction algorithm than on direct perception of mental effort. According to pupil data on dual-task, our results suggest a work-effort distribution between both tasks. Our interpretation is that, when carrying out two differentiation tasks at the same time, two or more brain areas are activated, and our limited attentional resources have to share them. This interpretation is closer to the results of Just et al. (2001), indicating that the activation of each brain area is lower under dual-task compared with the single-task condition. In this case, we suggest that the pupil could reflect the magnitude of the maximum activation instead of the sum of the activated areas.

The following argument also supports the former hypothesis. We have seen that the more cognitively workload tasks appear associated with higher visual detection impairment. Assuming that the participants attempt to optimize performance of both tasks (which would be in accordance with the reduced activation in both brain areas by distributing a maximum of mental effort), one could expect that the observed visual detection impairment would be also accompanied by cognitive task performance impairment. A complementary analysis of the performance of the two more workload tasks, talking and calculating,

revealed the existence of such impairment. This allows us to infer the hypothesis of shared resources between tasks, and that the total activation may not be higher than when tasks are performed alone, either according to pupil data or when considering the task performance decrement.

With regard to blink rate, the hypothesis about blink rate is more complex than that of NASA-TLX or pupil, because it includes different expectations depending on the distinction between visual demand (visual workload) and mental workload. According to our expectations, the higher visual demand of detecting vs. no detecting would lead to a higher blink-rate inhibition effect for detecting. In contrast, cognitive tasks with higher mental workload, such as talking or calculating, would produce a higher blink rate when compared to listening or the control conditions. The results support the formulated hypothesis. Blink rate effects allow us to distinguish between visual demand and mental (no visual) workload. If the participant is asked to look at a screen with changing visual stimuli and detect a target letter, we find a significant blink-rate reduction when comparing with no detecting conditions. We believe that this blink-inhibition process needs attentional resources. But if subjects have to use these attentional resources in another mental task, they will remove it from the visual search, and, in consequence, the blink rate increases. Therefore, our interpretation is that cognitive tasks interfere with the blink inhibition process due to the visual search.

Regarding correlations, and taking applied goals into account, the proportion of correct detection responses (hit rate) is the most relevant objective criterion to evaluate the impairment of a visual search attributable to simultaneous performance of a cognitive task. For dual-task, in the figures, it can be seen that the higher the workload of the cognitive task, the more impairment in detection. So, we think that mental workload measures could be good predictors of visual performance, and that such measures can be taken either in dual-task or in single-task conditions in order to determine the best predictor. For the single-task condition, when cognitive tasks are performed alone, correlations of mental load measures with hit rate would present a great applied value, that is: the effect of the cognitive task on a visual task could be predicted before performing that visual task. The highest correlation with detection occurred with the pupil-size effect when the mental workload is evaluated in single-task condition (-.69). The most reasonable interpretation is that detection performance decreases with increased mental workload and that the pupil effect under single-task is a more genuine evaluation of mental workload. As we have seen, the pupil alters its pattern when shifting from single- to dual-task. The performance of dual-task probably involves additional, not well understood, factors such as the coordination between tasks, causing undesirable biases in the measures of mental workload. Considering the mental load measures taken under dual-task conditions, NASA-TLX, compared with the other measures, appears to

be the best predictor of visual detection impairment (-.62) and thus, if the evaluation has to be made under dual-task conditions, this measure reflects better the complexity of the visual and mental workload involved.

The results of the present study contribute to reinforce some applied results obtained in real driving conditions (Recarte & Nunes, 2003), when considering the interference of listening versus producing verbal messages in a driver's visual processing. Now, we can state that those results were not due to eventually uncontrolled situations from the natural environment, but instead they represent a genuine reproducible interference process.

This result underlines the importance of interference between tasks of a different nature, which, a priori, could be thought not to share common resources, such as the case of visual processing impairment caused by purely cognitive tasks. When referring to "purely cognitive tasks," we want to note their nonvisual nature, given that there is an important research field, under the name of *inattention blindness* (Lavie, 2006; Mack & Rock, 1998; Most, Scholl, Clifford, & Simons, 2005), which focuses on attentional visual impairment effects, but this is always produced by another visual task which captures attention and causes "blindness." Now, we can propose a similar effect when using a cognitive task.

This has applications in driving. Mental workload measures are useful to evaluate whether certain cognitive activities, with no visual workload, may interfere with the driver's visual processing (i.e., the effect of speech on our detection of a deceleration manoeuvre of the vehicle in front of us). A previous evaluation of the pupil-size effect of the speech task performed alone would be the best predictor of visual interference.

## References

- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91, 276-292.
- De Waard, D. (1996). *The measurement of drivers' mental workload*. Unpublished doctoral dissertation, University of Groningen, Traffic Research Centre. Haren, The Netherlands
- De Waard, D. (2002). Mental workload. In R. Fuller & J. A. Santos (Eds.), *Human factors for highway engineers* (pp. 161-176). Oxford, Pergamon.
- De Waard, D., & Brookhuis, K. A. (1997). On the measurement of driver mental workload. In J. A. Rothengatter & E. Carbonell (Eds.), *Traffic and transport psychology* (pp. 161-173). Amsterdam: Elsevier.
- Fogarty, Ch., & Stern, J. A. (1989). Eye movements and blinks: Their relationship to higher cognitive processes. *International Journal of Psychophysiology*, 8, 35-42.
- Granholm, E., & Steinhauer, S. R. (2004). Introduction: Pupillometric measures of cognitive and emotional processing. *International Journal of Psychophysiology*, 52, 1-6.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.
- HASTE. (2002). *Human Machine Interface and the Safety of Traffic in Europe*. Project GRD1/2000/25361 S12.319626.
- Hoecks, B., & Levelt, W. J. M. (1993). Pupillary dilatation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments, & Computers*, 25, 16-26.
- Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004). Task-evoked pupillary response to mental workload in human-computer interaction (pp. 1477- 1480). In *CHI'04*. New York: AMC Press.
- Janisse, M. P. (1977). *Pupillometry: The psychology of the pupillary response*. New York: Wiley.
- Just, M. A., Carpenter, P., Keller, T., Emery, L., Zajac, H., & Thulborn, K. (2001). Interdependence of nonoverlapping cortical systems in dual cognitive tasks. *NeuroImage* 14, 417-426.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Kubose, T., Bock, K., Dell, G. S., Garnsey, S. M, Kramer, A. F., & Mayhugh, J. (2006). The effects of speech production and speech comprehension on simulated driving. *Performance Applied Cognitive Psychology*, 20(1), 43-63.
- Lavie, N. (2006). The role of perceptual load in visual awareness. *Brain Research*, 1080, 91-100.
- Mack, A., & Rock, I. (1998). *Inattention blindness*. Cambridge, MA: MIT Press.
- Most, S. B., Scholl, B. J., Clifford, E. R., & Simons, D. J. (2005). What you see is what you set: Sustained inattention blindness and the capture of awareness. *Psychological Review*, 112(1), 217-242.
- Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.
- Nunes, L. M., & Recarte, M. A. (2004). Speed, traffic complexity, and visual performance: A study on open road. In G. Underwood (Ed.), *Traffic and transport psychology: Theory and application* (pp. 339-354). Amsterdam: Elsevier.
- O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 42/1-42/49). New York: Wiley.
- Recarte, M. A., & Nunes, L. M. (2000). Effects of verbal and spatial-imagery task on eye fixations while driving. *Journal of Experimental Psychology: Applied*, 6, 31-43.
- Recarte, M. A., & Nunes, L. M. (2002a). Parpadeo durante la conducción: efectos de la carga mental y del tiempo conduciendo. *Vigilia-Sueño*, 14 ( Supl.), 161-167.
- Recarte, M. A., & Nunes, L. M. (2002b). Mental load and loss of control over speed in real driving: Towards a theory of attentional speed control. *Transportation Research, Part F* 5, 111-122.
- Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision

- making. *Journal of Experimental Psychology: Applied*, 9, 119-137.
- Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX and workload profile methods. *International Review of Applied Psychology*, 53, 61-86.
- Ryu, K., & Myung, R. (2005) Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics*, 35(11), 991-1009.
- Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology*, 52, 77-86.
- Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink rate: A possible measure of fatigue. *Human Factors*, 36, 285-297.
- Siveraag, E. J., & Stern, J. A. (2000). Ocular measures of fatigue and cognitive factors. In R. W. Barks & W. Boucsein (Eds.), *Engineering psychophysiology: Issues and applications* (pp. 269-287). Mahwah, NJ: Erlbaum.
- Vicente, K. J., Thornton, D. C., & Moray, N. (1987). Spectral analysis of sinus arrhythmia: A measure of mental effort. *Human Factors*, 29, 171-182.
- Wickens, C. D. (1992). *Engineering psychology and human performance*. New York: HarperCollins.

Received April 18, 2008

Revision received June 5, 2008

Accepted July 7, 2008